Check for updates

Journal of
APPLIED PHILOSOPHY

Journal of Applied Philosophy, Vol. 41, No. 1, February 2024
doi: 10.1111/japp.12710

How Does Human Agency Actually Work? On Bratman's 'Core Capacity Thesis' and the Relation between Philosophy of Action and the Empirical Sciences

JONATHAN PHILLIPS AND DAVID PLUNKETT

ABSTRACT *Throughout his career, Michael Bratman has developed a detailed model of individual 'planning agency', and, more recently, models of joint action and aspects of social life that he argues such planning agency helps support. How might we empirically investigate whether these models capture what is going on in actual human lives? In this article, we critically engage with this broad question by focusing on what Bratman calls the 'core capacity thesis', which is at the center of his most recent discussions of how his work in the philosophy of action hangs together. In his own words, the core capacity thesis claims that 'our capacity for planning agency plays basic roles in multiple, interrelated systems of human, mind-shaped practical organization, individual and social'. We explore how one might go about testing this thesis, focusing on Bratman's particular way of defending this thesis using the idea of 'nested Cummins functions'. In so doing, we hope to contribute to a broader discussion about what the relationship is (or, perhaps more importantly, should be) between the kind of philosophy of action Bratman does and empirical work in fields such as psychology, cognitive science, and biology.*

1. Introduction

What do philosophers aim to accomplish in the philosophy of action? If we ask this question at a high level of generality, any adequate answer to it is going to have to make room for some significant variation. For example, consider the different aims (not to mention the methodological approaches) of Elizabeth Anscombe, Donald Davidson, and Christine Korsgaard in their work on action, agency, and related topics.[1] In this article, we want to focus on one important aim that runs through the philosophy of action, which is tied to significant portions of the subfield. In short, the aim is to help illuminate how human agency actually works. This aim, in turn, might then be part of a broader one: namely, to illuminate how a wide array of socially rich forms of human interaction tied to human agency actually work.

In "Planning and Its Function in our Lives", Michael Bratman argues that his work in the philosophy of action, which he has now pursued for over four decades, is fruitfully interpreted as involving this aim.[2] In much of his groundbreaking work in this area, however, Bratman does not approach this aim head-on. Instead, he has proceeded in a stepwise way. In the core of his work, Bratman aims to build models of how planning agency (and then later how distinctive forms of social life) might be realized. The idea is that these models might, in turn, capture what is going on in key parts of actual human lives.

Here is a rough gloss on how this approach is meant to work. Over the course of his career, Bratman has developed an account of a distinctive kind of agency that he calls

© 2024 Society for Applied Philosophy

'planning agency'. Put roughly, for Bratman, planning agency involves the ability to form intentions, which are mental states that allow planning agents to guide their activity over time in distinctive ways. According to Bratman, planning agency also makes possible certain distinctive ways that agents can coordinate with each other, in both small-scale groups and larger ones. Bratman provides detailed models of how planning agency might be realized at the individual level, and how such individual planning agents might then interact in the right ways to engage in such things as joint action, norm-guided social behavior, and the creating and running of social institutions.[3]

In much of his work, Bratman is content to provide 'sufficiency conditions' for the phenomena he is interested in.[4] That is: he aims for his models to identify sufficient conditions for realizing the phenomena at hand (e.g. joint action among planning agents), rather than providing (more ambitiously) necessary conditions for doing so, or (even more ambitiously) a real definition of those phenomena or an account of their essence. If successful, his models would show one possible way for agents to be planning agents, and then, in turn, one possible way for such agents to interact to make possible forms of joint action, norm-guided social behavior, and complex, large-scale social institutions.[5] It is a separate question whether his models apply to particular actual agents, in the actual world. For example, we can ask: do actual human beings realize planning agency (and the further things it purportedly makes possible) in the way Bratman proposes in his models? What about chimpanzees or dolphins? Bratman (quite sensibly) thinks that armchair philosophical reflection alone will only get us so far in answering these questions, and that comprehensive answers will require the kind of empirical inquiry that scientists specialize in. In other words, to answer these questions, Bratman thinks that we will need to look to empirical results from scientific inquiry (e.g. from empirical psychology, cognitive science, biology, and the like).

Here is an important question that then emerges from this: what kind of empirical inquiry would help us answer questions about whether Bratman's models correctly describe actual human beings in the actual world? This is the broad question we take up in this article. In short, we consider which results from scientific inquiry could help us figure out whether Bratman's models accurately describe our lives as humans (e.g. our forms of agency and social life). Sorting out this question, we think, matters for multiple reasons. First, getting clearer on what empirical evidence matters here can help us better understand the content of Bratman's models: for example, what they are in fact committed to with regard to planning agency, and how the different components of his models are meant to fit together. Second, it can help us better understand what kind of relationship the kind of philosophy of action Bratman is engaged in has (or perhaps *should* have) to the empirical sciences. In discussing these issues, our aim is not to make an argument either for or against Bratman's views in 'Planning and Its Function in our Lives' or elsewhere. Rather, it is to raise a series of questions that, we hope, prompt Bratman to further clarify his overall views in the philosophy of action (and related areas), and, more generally, help all of us get clearer on the costs and benefits of different ways of thinking about the broader issues at hand.

2. The Core Capacity Thesis and 'Nested Cummins Functions'

Bratman's models of individual and shared planning agency – models that he develops to provide sufficient conditions for realizing the phenomena at hand – are detailed ones, full of nuanced and intricate claims. If we are wondering about the degree to which these

© 2024 Society for Applied Philosophy

models help illuminate our actual lives as human beings, trying to assess the accuracy of the entire models all at once would be a fairly daunting task. One more manageable way of approaching the overall question is to start by taking on different parts of the models – such as different theses, claims, etc. within them – and trying to determine how much they illuminate our lives as human beings. There may, of course, be methodological limits to how much we can separate out different parts of the models from each other to assess their accuracy or explanatory power independently, given the holistic nature of such assessment in general. But even if that is the case, that does not mean it is a mistake to try to start in small steps to begin such overall holistic assessment.

With this in mind, in this article, we approach our overall, guiding question that we introduced above – namely, the question of what kind of empirical inquiry we could use to test whether Bratman's models apply to us as human beings – by focusing on the main thesis at the heart of 'Planning and Its Function in our Lives', which he calls the 'core capacity thesis'. Put roughly, this thesis states that 'our capacity for planning agency plays basic roles in multiple, interrelated systems of human, mind-shaped practical organization, individual and social'.[6] We focus on this thesis for three reasons. First, Bratman takes it to be an important part of a new way of thinking about his work in the philosophy of action, which helps 'deepen our understanding of the planning theory of human agency and its significance'.[7] Second, Bratman explicitly takes it to be a 'conjecture'[8] about human beings that is 'subject to empirical study'.[9] Third, although (as we explain below) it is not entirely clear which aspects of Bratman's overall models the thesis incorporates, we take it that it is not meant to just be a new label for the idea that all of Bratman's models, in their full detail, accurately describe the relevant parts of human life.

What, more exactly then, is the core capacity thesis? The thesis claims that planning agency is a 'core capacity' of human beings that helps explain what is going on in human life at the different levels that Bratman's previous work has focused on, ranging in scale from the way in which an individual organizes her life over time to how we form (and run) social institutions. As he puts it, 'our human capacity for planning agency is, for us, a core capacity, one that underlies multiple, interrelated forms of mind-shaped practical organization that play fundamental roles in our lives'.[10] For the reasons that we just described above, Bratman takes this thesis to be answerable to empirical inquiry about how actual humans actually operate. He writes: 'The conjecture, subject to empirical study, is that our human temporally extended, social and institutionally shaped practical lives significantly involve such planning agency and that our capacity for such planning agency significantly underlies our capacities for these forms of practical organization. This is the *core capacity thesis*'.[11]

In order to flesh out and defend the core capacity thesis, Bratman appeals to a particular kind of function (what Peter Godfrey-Smith calls a 'Cummins function', named after the work of Robert Cummins).[12] As Bratman puts it: 'what is primarily to be explained by appeal to a Cummins function of X is a capacity of a larger system within which X is embedded'.[13] Consider the human heart. Its circulating blood is a 'Cummins function' of it, 'insofar as the heart's doing this plays a role in the explanation of a relevant capacity of a relevant, larger system – in this case, the capacity of a human to stay alive'.[14] As this example illustrates, Cummins functions are relative to a particular larger system (e.g. the human body), and to particular aspects of what it is doing (e.g. staying alive). As Bratman underscores, in cases of central theoretical interest, the Cummins functions we are after are tied to larger systems that are 'important' in some sense, whether for

© 2024 Society for Applied Philosophy

our practical or theoretical projects, or both. So too with which aspects of those systems we are focused on.[15]

Bratman argues that core parts of his work in the philosophy of action can be understood as a series of connected claims about Cummins functions. Put roughly, his idea is that, first, we start with a functionalist account of what 'intentions' and other mental states are that play a role in planning agency. Then, as we move up in scale from individual human agency to the composition and activity of large-scale social institutions, we can 'construct' different forms of human agency and social organization by making a series of claims about nested Cummins functions. More specifically, he argues that 'at each step – beginning with individual temporally extended planning agency and then proceeding sequentially to each of the further constructions along the way – the constructed social psychological structure at that step has a Cummins function of supporting the practical organization highlighted in the next steps of the sequence. This is a system of *nested Cummins functions*.[16] For example: his work on small-scale joint action (e.g. two agents taking a walk together) can be seen as using resources from his account of individual agency to show how such joint action may be made possible. In turn, such small-scale joint action supports larger-scale social interactions, such as agents engaging in large-scale social practices (including of the kind Bratman takes to be illuminated by a version of H.L.A. Hart's account of 'social rules').[17] And then, in turn, such social practices support the constitution and running of large-scale social institutions (such as universities, corporations, or legal systems). The idea, in short, is that the various steps along the way – each articulated in rich detail in Bratman's prior work – are part of a general story about how to explain the relevant phenomena at each level in terms of a system of nested 'Cummins functions'.

Given this explanatory goal, how might we go about studying whether Bratman's core capacity thesis is correct? To answer this question, we think we need more clarity on what explanation the 'core capacity thesis' as such (as opposed to other parts of Bratman's theories) is in fact committed to regarding the relevant phenomena in human life.

3. How Many of Bratman's Philosophical Views Are Built into the Core Capacity Thesis?

The first issue we want to start with concerns how many of Bratman's philosophical views are built into the core capacity thesis.

One question here is whether the core capacity thesis *as such* is committed to the idea of Cummins-style functional explanations. Consider again Bratman's canonical formulation of the core capacity thesis that we introduced above, which states that 'our human temporally extended, social and institutionally shaped practical lives significantly involve such planning agency and that our capacity for such planning agency significantly underlies our capacities for these forms of practical organization'.[18] That statement by itself does not mention Cummins-style functional explanations. So, given this statement, it seems conceptually possible that the core capacity thesis could turn out to be true, but not because of the accuracy of certain Cummins-style functional explanations of the kind that Bratman focuses on (or, indeed, functional explanations of any kind). For example, as stated, it seems that (in theory) one might be able to defend the core capacity thesis through a series of either causal claims that are not put in the relevant Cummins-style

© 2024 Society for Applied Philosophy

functional terms or metaphysical ones (e.g. claims about metaphysical grounding, or other forms of asymmetric metaphysical dependence), or some combination of these. This suggests that Bratman's embrace of the idea of 'nested Cummins functions' is his *preferred way* of making good on the core capacity thesis, rather than something built into the thesis itself. On another way of reading Bratman, however, the core capacity thesis itself is committed to the idea that 'nested Cummins functions' (of some kind or another) are the relevant way that planning agency is involved in, or underlies, the relevant phenomena at a larger scale. This way of reading Bratman is arguably suggested, for example, by one reading of Bratman's statement that '[t]he primary focus of the core capacity thesis is directly on the idea that our capacity for planning agency has the Cummins function of supporting interrelated forms of mind-shaped practical organization that are fundamental to our human lives'.[19] We are not sure of the best way to read Bratman's considered view, with respect to whether or not the core capacity thesis as such involves a commitment to Cummins-style functional explanations.

Suppose we focus on ways of attempting to vindicate the core capacity thesis that (whether optionally, or by necessity) proceed via the idea of 'nested Cummins functions', as Bratman does. Even with that framework in hand, we still have many open questions about how much the core capacity thesis is committed to, with regard to the rest of Bratman's theories in the philosophy of action (and related areas of philosophy). Consider three different salient options about how best to understand the commitments of the core capacity thesis on this front:

> Option 1: The core capacity thesis is neutral on whether individual human planning agency is best understood along the lines of Bratman's theory of planning agency or not. All that matters is that planning agency (of some kind or another) is a core capacity in the relevant sense. Furthermore, the core capacity thesis is neutral on what the 'nested Cummins functions' involve beyond the individual level of planning agency.

> Option 2: The core capacity thesis endorses the idea that individual human planning agency *as understood along the lines of Bratman's theory of planning agency* is a core capacity in the relevant sense. But it is neutral on what the 'nested Cummins functions' involve beyond the individual level of planning agency.

> Option 3: The core capacity thesis endorses the idea that individual human planning agency *as understood along the lines of Bratman's theory of planning agency* is a core capacity in the relevant sense. Furthermore, the core capacity thesis takes a stand on what the 'nested Cummins functions' involve beyond the individual level of planning agency. In particular, it involves endorsing *Bratman's developed theories* (e.g. of shared action and institutional activity) as ways of explaining the relevant phenomena at each different stage of the relevant 'construction' in human life.

These are far from the only options worth thinking about. For example: one might opt for something between Option 2 and Option 3, on which the core capacity thesis is not committed to *all* of the details of Bratman's theories beyond the level of individual planning but is committed to *some* of those details. There are many such options in that space, depending on which aspects of Bratman's theories one focuses on. We put forward the

© 2024 Society for Applied Philosophy

three options not because we think that one of them is the most plausible interpretation of Bratman's views, but rather to help illustrate the range of possibilities here.

For our purposes, a key point is that these options involve different theoretical commitments, which would require different forms of empirical support to establish. To press this point, consider again Bratman's 'strategy of sufficiency'. It could well turn out that, even if Bratman successfully provides sufficiency conditions for the relevant phenomena at each stage of his overall account, some other theory could *also* provide sufficient conditions for the instantiation of the phenomena at hand. This possibility (which Bratman takes seriously throughout his work) underscores some of the reasons why it matters how many of Bratman's overall theories are built into the core capacity thesis as such.

For the core capacity thesis to be an interesting *part* of Bratman's philosophical views about the topics at hand – and not, for example, just a summary of what those views are – we think that the core capacity thesis (assuming, for now, that one aims to vindicate it via the strategy of 'nested Cummins functions') must be understood as something along the lines of either Option 1 or Option 2 above, rather than Option 3, or perhaps something between Option 2 and Option 3. But we also are not entirely sure which of these options (or some other one that we have not discussed) Bratman thinks best captures what he has in mind with the core capacity thesis.

4. What Is the Scope of the Core Capacity Thesis?

Now consider the issue of which phenomena in human life Bratman takes the core capacity thesis to concern. In one sense, Bratman is fairly clear about this question: namely, he wants to explain many important aspects of human life, ranging from aspects of our individual lives to aspects of our larger forms of shared activity. But which aspects more exactly? And how many of those relevant aspects need to be explained in the way that the core capacity thesis claims they are? Suppose, for the sake of argument, we were to discover that *some* of our shared activities, social institutions, etc. are organized in the ways that Bratman's models describe, but that *some other* ones diverge from Bratman's models? If many of those phenomena turn out to be both 'important' (in the sense that matters to the core capacity thesis), and best explained in other ways, how much of a problem would that be for the core capacity thesis?

Suppose we understand the core capacity thesis along the lines of Option 2 above (which, recall, incorporates the idea that individual human planning agency *as understood along the lines of Bratman's theory of planning agency* is a core capacity in the relevant sense, but is neutral on what the 'nested Cummins functions' involve beyond the individual level of planning agency). Then, if it turns out that Bratman-style planning agency does not underlie a wide range of the important relevant phenomena in human life, that would seem to put pressure on the core capacity thesis.[20] Relatedly, suppose it turns out that Bratman-style planning agency does play a role in the best accounts of the relevant phenomena, but not in the way that Bratman thinks it does in his more developed models of how this works. This result would seem to put pressure on Bratman's particular way of fleshing out and defending the core capacity thesis, including his particular account of what the nested Cummins functions are, and how they work (or it would directly put pressure on the core capacity thesis itself, if it were understood along the lines of Option 3 above, which, recall, involves all of Bratman's developed theories at each stage of the

© 2024 Society for Applied Philosophy

'construction' of the relevant human phenomena). This possibility brings out the following issue: *which* and *how many* relevant phenomena of intra- and inter-personal practical organization does Bratman's account need to explain for the core capacity thesis to be vindicated?

Here it is worth noting that, in much of his work, Bratman is tentative about which aspects of human activity align with his models. For example, in his most recent book, *Shared and Institutional Agency*, Bratman explores the idea that our legal institutions work in the way his model describes.[21] To do so, he draws on Hart's account of 'social rules', which was at the core of Hart's account of legal institutions.[22] Importantly, however, Bratman does not commit to the idea that his model in fact explains legal institutions, let alone the idea that it would be a problem for the core capacity thesis if his model did not explain those institutions in particular.

What this brings out is this: there are open questions about the *scope* of the core capacity thesis, in terms of which aspects of human sociality (or which social institutions, etc.) the thesis is meant to cover. This open-endedness, in turn, makes it unclear what kind of examples would either help confirm the thesis or else be counterexamples to it.

In response, Bratman might well insist that he is allowed to be vague and somewhat open-ended here. He can insist that his account of nested Cummins functions captures many important parts of human activity (including important parts of our sociality, our institutions, etc.) and just leave it at that. In one sense, that is fine. After all, he is giving us general models, which might well help us understand many of the relevant human phenomena, even if not all of them, nor even such-and-such specific ones. The more aspects of our lives these models help illuminate, the better. Moreover, these models might help illuminate aspects of our lives whether they describe those with total accuracy or just with partial accuracy.

We think these points are all relevant for assessing the overall import of Bratman's models. Our point here is just that, as stated earlier, there is a lot of room for variation in how exactly we should understand the core capacity thesis, and thus also what empirical information we should be looking for in evaluating *it* in particular. Tied to this point, this issue about scope also matters for how much significance we should give to the moral and political considerations that he raises at the end of his article, stemming from his core capacity thesis, about the importance of fostering the right sorts of environments to allow people to realize their capacity for planning agency in robust ways. Those remarks are (we think) charitably understood as premised on the idea that Bratman-style planning agency underwrites large (morally and politically important) swaths of our social lives.

5. How Could We Empirically Test a Model of Nested Cummins Functions?

Suppose we settle on answers to the last two sets of issues we raised, such that we have a clearer sense both of how many of Bratman's overall views are built into the core capacity thesis, and also of which aspects of human life the core capacity thesis concerns. We then face a further, and, we think, deeper issue: it is not clear which direct empirical predictions the core capacity thesis makes by positing a system of nested Cummins functions. In other words, we need to resolve how one could empirically test an explanation of the relevant phenomena in terms of a system of nested Cummins functions along the lines that

© 2024 Society for Applied Philosophy

Bratman proposes. There are different ways of interpreting Bratman's proposed idea of nested Cummins functions that vary in their level of empirical commitment. For ease of exposition, we will work with what we called Option 2 above in Section 3: namely, the idea that the core capacity thesis proposes that individual human planning agency *as understood along the lines of Bratman's theory of planning agency* is a core capacity, and the base of a system of nested Cummins functions that explain the relevant further forms of human social interaction, sociality, and institutional activity, *in one way or another* (whether following the details of Bratman's own proposed models here, or not).

One way of attempting to make tractable the empirical predictions of Bratman's proposed nested Cummins functions would be to understand these claims as analogous to nested functions in computer programming, where one function 'calls' another function, meaning that the execution of the former function involves (among other things) the execution of the latter function. To illustrate, you may have a function that takes images from a security camera as input and returns an alert whenever a door opens, and this function may call another function that is designed to identify the door in each image. Applying this way of understanding nested functions to Bratman's core capacity thesis allows us to precisify the empirical predictions: the causal process that instantiates our capacity for joint action would involve the more basic causal process that instantiates our capacity for diachronic planning within our own individual lives, via the formation of Bratman-style intentions. Yet-higher-level capacities, such as large-scale social organization, could then be understood as functions that recruit functions (e.g. those that explain joint action) that recruit functions (e.g. those that explain individual intentions), allowing us to recover Bratman's proposed nested hierarchical structure.

While we are not sure whether Bratman would endorse this way of understanding the core capacity thesis, this approach has the benefit of being able to make relatively clear empirical predictions because it clearly specifies the dependence relation between particular causal processes. For example, if we discovered that the human brain causally instantiates the function of diachronic planning in some specific set of brain regions involving, for example, the prefrontal cortex (PFC), we would then expect that same causal process to be involved when people engage in joint action or any other higher-level function in which this capacity is nested. In other words, because the human brain may causally instantiate some particular function in a signature way, we could test, perhaps using fMRI (functional Magnetic Resonance Imaging), whether *that same function* was causally involved in some higher-level function in which it was proposed to be nested by looking for the same signature patterns of brain activity (e.g. in the PFC). This way of clarifying the empirical predictions of the core capacity thesis also has the virtue of roughly aligning with how cognitive scientists have thought of other kinds of 'core cognition'.[23] For example, Susan Carey has argued that a 'core' system for precisely representing small numbers ($\sim$1–4) is recruited by the subsequently-developing capacity for learning number concepts, which is then used for representing larger numbers.[24]

The ideas we have just floated involve attaching claims about Cummins functions to causal claims of some kind. We take this process to be in the spirit of what Bratman himself endorses about Cummins functions. For example, when introducing the basic idea of such functions, he claims that an appeal to a Cummins function is 'appeal to a specific kind of causal-functional property'.[25]

But perhaps the core capacity thesis, as cashed out in terms of a system of nested Cummins functions, does not involve a commitment to causal claims at the level of

© 2024 Society for Applied Philosophy

specificity we have just been discussing, even if it does involve appeal to causal-functional properties in some way. An alternative that is open to Bratman would be to argue that the core capacity thesis is not committed to the exact same causal process being recruited by each higher-level capacity in which it is nested. Instead, he may argue that the core capacity thesis was meant to pick out a *kind* of causal process, and then propose that this *kind* of causal process is involved in the function of a higher-level capacity (e.g. the capacity for joint action involves the kind of causal process that instantiates diachronic planning), and that this *kind* of process may be realized in many different ways. This leaves open the possibility that the same causal process may be physically instantiated in different ways across different cases. This approach thus need not predict, for example, that the neural processes recruited for diachronic planning will also be recruited in cases of joint action, or other higher-level capacities. While this approach leaves open the matter of how to delineate kinds of causal processes, it would still seem to make some testable empirical predictions. For example, it plausibly makes predictions about the stages of human development. This is because it would be surprising to find a point in human development in which there is the capacity for some higher-level ability (e.g. joint action) without the capacity for some lower-level ability (e.g. individual diachronic planning within one's own life), since (by hypothesis) the higher-level capacity would recruit the same *kind* of causal process as the lower-level capacity.

While these ways of understanding the core capacity thesis have the benefit of making relatively clear empirical predictions, those predictions may not be plausible. For example, we do not know of any positive evidence that would suggest large-scale shared actions of the kind involved in social organizations (on Bratman's theory of them) recruit the same brain regions involved in diachronic planning (of the kind that Bratman is focused on in his work on individual planning agency). Moreover, young children engage in joint action (or at least things that many psychologists identify as 'joint action') quite early in development, long before they seem to succeed in anything like longer-term diachronic planning of the kind meant to be afforded by Bratman-style intentions.[26] For example, simple forms of joint action (e.g. coordinated social exchanges) can be found as early as a few weeks after birth, and joint action involving sustained games involving jointly-attended objects emerges around 18 months, at which point such joint action is often initiated by the infant.[27] In contrast, even very simple forms of diachronic planning (involving coordinating motor actions on very short timescales) do not seem to emerge until late in the second year of life. For example, 21- and some 27-month-old children exhibited remarkable difficulty in achieving a simple goal when it involved planning over multiple sequential steps (e.g. putting together two halves of a plastic barrel with a small wooden block inside to make a rattle).[28] In fact, most developmental psychologists pin the emergence of genuine diachronic planning substantially later than this, well after three years of age.[29]

Of course, in both kinds of empirical predictions discussed above, there is still much room for debate. One could argue, for example, that the correct test for whether large-scale social activity involves the same causal processes as individual diachronic planning has simply not yet been done. Or one could push back against the developmental claim by attempting to clarify the standard for what is required for some behavior to count as genuine joint action, and then argue that young children do not in fact engage in joint action before their ability for individual diachronic planning (over the relevant timescale) emerges. Additionally, one may try to clarify the criteria for what counts as the ability for diachronic planning of the relevant sort, and argue that some form of the ability can be

© 2024 Society for Applied Philosophy

observed earlier in development than joint action, even if that ability is not as robustly realized as it is in later life. Each of these responses would then necessitate further empirical and theoretical work to settle the issue, but for now, we will simply note these possibilities as an important avenue for future work.

Setting these approaches to one side, a yet weaker alternative, which would not be committed to empirical predictions of the kind we have been discussing so far, would be to argue that Bratman's claim about nested Cummins functions should not be understood as picking out specific causal processes (or even specific kinds of causal processes). Instead, one could see it as pointing out that one fruitful way of explaining a higher-level ability, such as joint action, would involve appeal to another kind of ability, such as individual diachronic planning. There are at least two different ways of unpacking this last kind of idea, which might be pursued separately or together. First, one might hold that Bratman's model of nested Cummins functions is a useful *interpretation* of the set of empirical facts that are discovered about the relevant phenomena, e.g. large-scale social organization and joint action. Second, one might hold that his model of nested Cummins functions helps provide some sort of *non-causal explanation* of the relevant phenomena. A clear downside of either approach is that they are less clear on how to vindicate Bratman's claim that the core capacity thesis is 'subject to empirical study'.[30] There are, to be sure, moves that Bratman might make here. For example, he might draw on ideas about theory-choice in the sciences that emphasize the importance of factors other than predictive power, or on different ideas about non-causal explanations within scientific theorizing. Our point is just that abandoning the idea of nested Cummins functions making causal claims of some kind involves significant theoretical burdens, including ones that do not all obviously mesh well with Bratman's more general theoretical commitments.

We are not sure which of the empirical predictions we have discussed, if any, Bratman intends to take on in the core capacity thesis. And we are not sure which (if any) he should. Our point in putting them forward is just to present some examples that bring out the kinds of predictions he might make, insofar as the core capacity thesis is cashed out and defended in terms of a system of nested Cummins functions. If the core capacity thesis really is answerable to empirical study – which it certainly seems like it should be, given the kind of thesis it is – it is critical to know more about which empirical predictions the thesis is making. In addition to (obviously) helping set the stage for future empirical work testing the applicability of Bratman's models of planning agency to human life, learning more about what those empirical predictions are can help us better understand what the core capacity thesis is (as well as what Bratman is committed to more broadly in the models he has developed within the philosophy of action).[31]

6. Conclusion

As Bratman has emphasized throughout his career, one way that philosophical model-building (of the kind that is at the core of his work) can earn its keep is by showing that it helps us make progress in other projects that we have reason to think are independently valuable. Not all such projects need to be focused on how to understand human beings. To illustrate, one way Bratman's models might turn out to be fecund is if they helped researchers build robots with advanced forms of artificial intelligence that successfully instantiated 'planning agency', and then (more ambitiously) the relevant forms of shared

© 2024 Society for Applied Philosophy

activity, sociality, and institutional activity that Bratman's models have focused on. If engineers were able to build such robots along Bratmanian lines, that would be concrete proof that his 'strategy of sufficiency' can be helpful. This kind of potential impact of Bratman's work is likely to become increasingly important as humans make strides in studying and building machines with artificial intelligence, including forms of it which might turn out to be quite alien to human intelligence in fundamental ways. It is worth noting here that Bratman's work has already had a significant impact on certain important strands of research in computer science.[32] From another angle, Bratman's work might also help us understand the lives of not only human, but also non-human animals, as we gestured to at the start of this article. Here we can note, as Bratman himself underscores, that some of his ideas about planning agency have influenced Michael Tomasello's work on (non-human) great apes, and how their minds and behavior are both related to and distinct from ours.[33]

Part of our idea in this article, however, is that just stating these kinds of impacts (whether now, or in the future) is not enough to answer the questions we might have about how to assess the connection of Bratman's work to the empirical sciences. When evaluating the impact of Bratman's models (or the models developed by any other philosopher, on any topic), it is important to keep in mind the difference between two distinct ways in which his sort of model-building work can pay off. First, his models might inspire the work of scientists, even if those scientists depart from his models in significant ways as they develop their own theories (which, for example, is what Tomasello does in his own engagement with Bratman's work). Second, it might be that scientific work of a certain kind *empirically vindicates* the claim that Bratman's models (whether fully, or just partially) correctly describe the relevant phenomena. It is this second kind of possibility that we have been pushing on in this article, where we have been focused on the case of human agency and the various forms of social life and social institutions it helps support, and on one particular thesis (the core capacity thesis) at the heart of Bratman's most recent work.

In discussing what it would take to empirically vindicate Bratman's core capacity thesis, we hope to have helped contribute to an ongoing discussion about what connections there are (or, perhaps more importantly, *should be*) between philosophy of action (of the kind Bratman has pursued) and the empirical sciences. In closing, we want to underscore that our discussion has only focused on one place among many where these different projects can connect. Our discussion brings out a number of salient places where someone might disagree with important claims in Bratman's overall project in the philosophy of action. Here are three such places: (a) what the 'design specification' is for the relevant phenomena at hand (e.g. what it would take to be a 'planning agent', engage in 'joint action', or be a 'social institution' in the relevant senses), (b) what are sufficient ways of realizing those phenomena, and (c) how these phenomena work in actual human lives. Our focus in this article has been on the third issue, and, in particular, the role that empirical research should play in thinking about it. But notice that empirical research – say, of the kind carried out by Tomasello – might impact Bratman's project in the two earlier 'stages' of his project as well, in a way that suggests that there is a dynamic, ongoing conversation between philosophical and scientific theorizing on the issues at hand. To illustrate, consider that, through doing empirical research, our understanding of what core phenomena matter might change, as well as our understanding of what the relevant 'design specifications' for those phenomena are. For example: we might learn new things about psychology that change our understanding of what planning agency of the relevant kind must

© 2024 Society for Applied Philosophy

involve.[34] So too might empirical research reveal ideas about how relevant phenomena could be realized. For example: we might uncover new facts about certain biological organisms (or certain forms of artificial intelligence) that give us new insights into ways that planning agency might be realized.

As we hope our discussion brings out, there are many rich ways that work in the philosophy of action can both inform, and be informed by, work in the empirical sciences. We take it Bratman would readily agree. Our discussion here only scratches the surface in terms of understanding what forms this dynamic relationship might take. Our hope is that future work in the philosophy of action continues to push on increasing our understanding of these methodological issues, and, more generally, to pursue ongoing, rich interaction with the empirical sciences.

Jonathan Phillips and David Plunkett, Dartmouth College, Hanover, NH, USA. *david.plunkett@dartmouth.edu*

Acknowledgements

Thanks to Michael Bratman, Avery Kolers, Tristram McPherson, Adrian Russian, Zoe Thierfelder, Gideon Yaffe, and Manuel Vargas for helpful discussion and feedback. Thanks also to the Wissenschaftskolleg zu Berlin for fellowship funding for one of us (Plunkett), which helped support our work on this article.

NOTES

1 See, for example, Anscombe, *Intention*; Davidson, *Essays on Actions and Events*; Korsgaard, *Self-Constitution*. For a helpful critical overview of contemporary philosophy of action, see Paul, *Philosophy of Action*.
2 Bratman, "Planning and Its Function."
3 For some of Bratman's key work here, see Bratman, *Intention, Plans, and Practical Reason*; *Faces of Intention*; and *Structures of Agency* (all focused on his account of individual planning agency); *Shared Agency* (focused on small-scale joint action); and *Shared and Institutional Agency* (focused on larger-scaler shared activity and social institutions).
4 As he emphasizes in Bratman, "Planning and Its Function," 3.
5 This approach is tied to Bratman's way of pursuing Paul Grice's strategy of 'creature construction'. See ibid., 3, drawing on Grice, "Method."
6 Bratman, "Planning and Its Function," 3.
7 Ibid., 1. Bratman also discusses this thesis at length in *Shared and Institutional Agency*.
8 Bratman, "Planning and Its Function," 1.
9 Ibid., 1.
10 Ibid., 1.
11 Ibid., 1.
12 Ibid., 2, drawing on Godfrey-Smith, *Complexity*; and Cummins, "Functional Analysis."
13 Bratman, "Planning and Its Function," 2. Bratman contrasts such functions with what Godfrey-Smith calls 'Wright functions', named after Larry Wright. (See Godfrey-Smith, *Complexity*, discussing Wright, "Functions.") As Bratman puts it: 'What is primarily to be explained by appeal to a Wright function of X is why X is there' (Bratman, "Planning and Its Function," 4).
14 Bratman, "Planning and Its Function," 2.
15 Ibid., 2–3.
16 Ibid., 5.
17 For Hart's own discussion of 'social rules', see Hart, *Concept of Law*.

© 2024 Society for Applied Philosophy

18 Bratman, "Planning and Its Function," 1.
19 Ibid., 7.
20 Ibid., 3.
21 Bratman, *Shared and Institutional Agency*. In pursuing the idea of using his work on planning agency to account for legal institutions, Bratman is working in the same theoretical space as Scott Shapiro, who, drawing extensively on Bratman's earlier work, pursued the same basic strategy in Shapiro, *Legality*. Part of the contribution of Bratman's *Shared and Institutional Agency* involves the subtle ways of both building on and departing from Shapiro's views about how best to pursue this strategy.
22 Hart, *Concept of Law*.
23 See, for example, Carey, *Origin*; Kinzler and Spelke, "Core Systems."
24 See Carey, *Origin*, chap. 4.
25 Bratman, "Planning and Its Function," 2.
26 See Tomasello *et al*., "Understanding."
27 For a review, see Brownell, "Early Developments."
28 Bauer *et al*., "Planning Ahead."
29 For a review, see McCormack and Atance, "Planning."
30 Bratman, "Planning and Its Function," 1.
31 For readers interested in further discussion of how to empirically test Bratman's system of nested Cummins functions, and how that question connects to broader questions about the content and theoretical ambitions of Bratman's models of planning agency more generally, see Manuel Vargas's piece in this journal issue (Vargas, "Functionalisms"), which takes up some overlapping concerns, but with a different focus.
32 See, for example, the impact of Bratman's work, including Bratman *et al*., "Plans," and Bratman, *Intention, Plans, and Practical Reason*, in developing 'belief-desire-intention' ('BDI') software models in computer science. See Rao and Georgeff, "BDI Agents," for further discussion.
33 See, for example, Tomasello, *Natural History*.
34 Bratman gestures to this possibility when he introduces how he understands the basic elements of his model of cognition and action of which 'intentions' are a part. He writes: 'We understand intention, planning, evaluation, desire, belief, acceptance, and associated forms of thinking as interrelated elements in an overall causal system that involves associated forms of rational guidance. Our characterization of this overall system begins with our commonsense psychology but is potentially adjusted, modified, explicated, and systematized in the light of various theoretical and empirical considerations' (Bratman, "Planning and Its Function," 2).

References

Anscombe, G. E. M. *Intention*. Cambridge: Harvard University Press, 1957/2000.

Bauer, Patricia J., Jennifer A. Schwade, Sandi Saeger Wewerka, and Kathleen Delaney. "Planning Ahead: Goal-Directed Problem Solving by 2-Year-Olds." *Developmental Psychology* 35, no. 5 (1999): 1321–37.

Bratman, Michael E. *Faces of Intention: Selected Essays on Intention and Agency. Cambridge Studies in Philosophy*. Cambridge: Cambridge University Press, 1999.

Bratman, Michael E. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press, 1987.

Bratman, Michael E. "Planning and Its Function in Our Lives." *Journal of Applied Philosophy* 41, no. 1 (2024): 1–15.

Bratman, Michael E. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press, 2014.

Bratman, Michael E. *Shared and Institutional Agency: Toward a Planning Theory of Human Practical Organization*. Oxford: Oxford University Press, 2022.

Bratman, Michael E. *Structures of Agency: Essays*. Oxford: Oxford University Press, 2007.

Bratman, Michael E., David J. Israel, and Martha E. Pollack. "Plans and Resource-Bounded Practical Reasoning." *Computational Intelligence* 4, no. 3 (1988): 349–355.

Brownell, Celia A. "Early Developments in Joint Action." *Review of Philosophy and Psychology* 2, no. 2 (2011): 193–211.

© 2024 Society for Applied Philosophy

Carey, Susan. *The Origin of Concepts*. Oxford: Oxford University Press, 2011.
Cummins, Robert E. "Functional Analysis." *Journal of Philosophy* 72 (1975): 741–764.
Davidson, Donald. *Essays on Actions and Events: Philosophical Essays, Volume 1*. Oxford: Oxford University Press, 1980/2001.
Godfrey-Smith, Peter. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press, 1998.
Grice, Paul. "Method in Philosophical Psychology (From the Banal to the Bizarre)." *Proceedings and Addresses of the American Philosophical Association* 48 (1974): 23–53.
Hart, H. L. A. *The Concept of Law*, 3rd ed. Oxford: Oxford University Press, 1961/2012.
Kinzler, Katherine D., and Elizabeth S. Spelke. "Core Systems in Human Cognition." *Progress in Brain Research* 164 (2007): 257–264.
Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.
McCormack, Teresa, and Cristina M. Atance. "Planning in Young Children: A Review and Synthesis." *Developmental Review* 31 (2011): 1–31.
Paul, Sarah K. *Philosophy of Action: A Contemporary Introduction*. Oxford: Routledge, 2020.
Rao, Anand S., and Michael P. Georgeff. "BDI Agents: From Theory to Practice." *Proceedings of the First International Conference on Multiagent Systems* 95 (1995): 312–19.
Shapiro, Scott. *Legality*. Cambridge: Harvard University Press, 2011.
Tomasello, Michael. *A Natural History of Human Morality*. Cambridge: Harvard University Press, 2016.
Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. "Understanding and Sharing Intentions: The Origins of Cultural Cognition." *Behavioral and Brain Sciences* 28 (2005): 675–735.
Vargas, Manuel. "Functionalisms and the Philosophy of Action." *Journal of Applied Philosophy* 41, no. 1 (2024): 41–55.
Wright, Larry. "Functions." *Philosophical Review* 82, no. 2 (1973): 139–168.

© 2024 Society for Applied Philosophy